I. Overview

This is the first of the monthly progress reports for the VIMSS Genomes To Life Project. Each core is expected to report on the goings on, technical, administrative and, if necessary, financial in their groups. I want to remind everyone how important to make sure everyone is communicating. The discussion boards (which are currently unused) provide a forum for people to ask questions about direction of the project, priorities, and technical issues that can be read and answered by the entire group. I know email is often the most efficient means but it does privatize some of the important communications. Also, posting to BioFiles is EXTREMELY important. This is the best metric I can give to the DOE leadership that we are making progress aside from the genomics websites. Please make us and yourselves visible by donating data and information to the website. Every PI should have at least one piece of substantive information in the database by next month.

II. AEMC

We have successfully maintained live cultures of *D. vulgaris* on solid and in liquid medium. We are currently up-scaling this process to maintain larger batch cultures of *D. vulgaris* in a sterile and anaerobic environment.

The iron content of the medium is being altered to range from trace amounts to 1 g/L to reduce the amount of precipitate formation in the medium without compromising the growth of the cells.

We have started to obtain baseline data of PLFA profiles of *D. vulgaris* at different stages of growth and on complex vs. defined medium. A growth curve will be generated for *D. vulgaris* on complex medium. These data will provide a baseline for comparison of *D. vulgaris* under oxygen stressed conditions (and future stressors). Initial PLFA analysis of *D. vulgaris* has been completed and is being duplicated. We are aiming for reproducible results and identification of unique *D. vulgaris* signatures.

We have also set-up some initial soil columns with *D. vulgaris* with Ken Williams in Jill Banfield's lab and plan to perform some assays on *D. vulgaris* harvested from these columns.

Fresh sediment cores from the 3 contaminated areas and the background area of the NABIR FRC were obtained by LBNL and shipped to UW and Diversa. These samples are being subjected to Biopanning™, TRFLP, and SRB (*Desulfovibrio*) isolation, PLFA, SRB probing, direct counts, enzyme assays, and enrichments.

SR-FTIR spectromicroscopy and SEM studies of the effect of atmospheric oxygen on *Desulfovibiro vulgaris*. [Hoi-Ying Holman at LBNL].

1) Obtain scanning electron microscopy (SEM) images of *D. vulgaris* exposed to oxygen. *D. vulgaris* which were grown anaerobically on agar plates were exposed to air and their SEM images were obtained for the following exposure time points: 5 minutes, 17 minutes, and 20 days. *Preliminary Results:* SEM images of *D. vulgaris*: We saw significant phenotype changes as cells were exposed to air. For example – cell enlargement, EPS formation and reduction of flagella. This appears to be a bacterial response to oxidative stress, I propose that these phenotype changes are associated with the formation of protective capsules as a stress response.
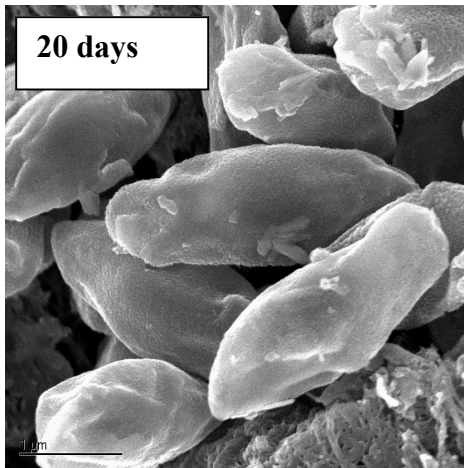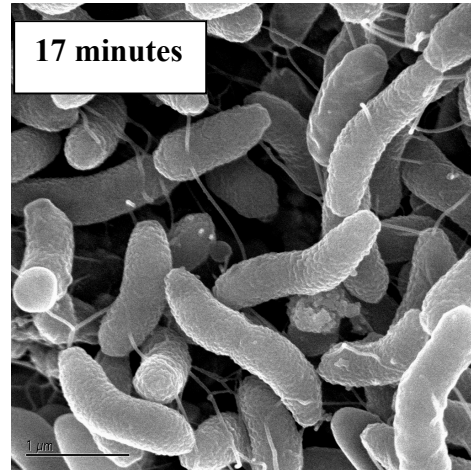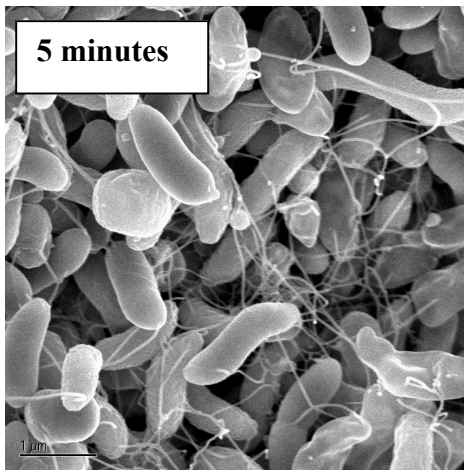
5 minutes

17 minutes

20 days

Figure 1. SEM images of *D. vulgaris* at three different exposure times.

2) Modify existing SR-FTIR spectromicroscopy apparatus to study *Desulfovibiro vulgaris* under anaerobic conditions. Re-design and construct the miniature anaerobic sample holder for the SR-FTIR microscope stage. Note: We are having difficulties with the optical aspects of the miniature anaerobic sample holder but hope to resolve this in April.

University of Washington

We received samples of subsurface sediments from non-contaminated and contaminated areas of FRC site on March 5, 2003. Sediment cores were recovered from FB609-10-00 (background area, core depth 180-239in), FB060-01-00 (area 1, depth 180-228 in), FB052-01-00 (area 2, depth 245-300 in) and FB056-01-00 (area 3, depth 194-198 in). The samples, shipped in plastic canisters filled with argon, were held at 4 C and transferred from the LBNL canisters to a new set of canisters (fabricated at the U of W) designed for manipulation in an anaerobic glove box and equipped with gassing ports to maintain anoxic conditions as needed. Each core was divided in to four equal depth intervals and subsamples from the deeper intervals used to inoculate MPN enrichments and evaluate alternative DNA extraction protocols. An additional subset of samples from each depth interval was transferred to50 ml plastic tubes for long-term storage at –80°C. Transfer and processing of all samples was conducted in plastic glove box under nitrogen.

MPN enrichments were set up for two subsamples on B2 medium reformulated to increase its buffering capacity and supplemented with one of the following substrates: 1) lactate, 2) acetate, 3) propionate and 4) hydrogen with carbon dioxide. No sulfate reduction has yet been observed.

Initial DNA extraction studies using the Mo Bio Kit with 5g of sediment did not yield DNA of sufficient quantity or quality for PCR amplification using general 16S rRNA primers. We are now investigating several alternative protocols that have been used successfully in our laboratory for DNA extraction from different environments.

Oak Ridge National Laboratory

We exchanged some information with UW collaborators for extraction of nucleic acids from FRC sediments. We have also started some SRB enrichments to be sent to Sergey. We hope to initiate some molecular analysis in the near future. We also did a test run for the isolation of biomass from approximately 100 liters of contaminated, FRC groundwater (FW-106 in Area 3). The cell number was estimated with AODC, and was approximately 110,000 cells/ml. A portion was used for DNA extraction from an agarose plug of cells, and another portion was used for a freezing-grinding method. Yields for both procedures were low.

III. FGC

Transcriptomics (ORNL)

Objectives:

       Compare oligonucleotide-based and cDNA-based arrays.

To perform initial microarray expression profiling studies for the model bacterium *S. oneidensis* MR-1 to establish a baseline response to various environmental stresses

Progress since last report:

We have done some initial experiments to compare oligonucleotide-based and cDNA-based arrays. Hybridization was performed at 42°C without formamide. Signal intensity increased with increasing length of the oligonucleotide probe, and the greatest intensity was observed with cDNA. The results suggested that cDNA arrays had increased sensitivity compared to oligonucleotide arrays in the presence of 50% formamide. However, 60mer and 70mer oligonucleotides had similar overall signal intensities without formamide to cDNA arrays with formamide. We have also evaluated currently available software and tested genome data from four organisms and three free programs, and then selected, modified or developed suitable software to design genome-wide oligonucleotide probes for *Desulfovibrio vulgaris* Hildenborough so that we can use the tool for further research on gene discovery and pathway identification. We recently used OligoPicker to design 386 oligonucleotides for selected ORFs.

Continuing with detailed analyses of microarray expression data from pH stress experiments comparing growth of MR-1 under neutral pH (pH 7) to growth under pH 4 or pH 10. Data has already been statistically analyzed using ArrayStat (see accompanying Excel file) and genes expressed specifically and exclusively in response to acidic pH or alkaline pH have been identified. A subset of genes differentially expressed in response to pH stress has been selected for further confirmation using real-time PCR. The writing of a manuscript describing this work is currently underway.

Microarray experiments analyzing the MR-1 transcriptome under heat shock conditions have been completed. In addition, an *rpoH* (sigma-32) mutant has been analyzed using whole-genome microarrays.

Studies looking at the response of MR-1 to metal toxicity are underway. These experiments are only in the growth-curve study stage in order to determine at what concentrations metals become toxic. At the moment, we are focusing on strontium and cobalt toxicity but plan to include technetium and perhaps chromium. Speciation analyses for strontium have already been completed.

Future Work

Future work includes the determination of specificity of oligo-probes at varying concentrations of formamide, as well as the results reported as ratios of up-regulated and down-regulated genes.

Future work includes printing the probes and testing the prototype array with labeled cDNA from *D. vulgaris* cells grown under different conditions.

Complete pH stress manuscript. Design more detailed time-series experiments.

Complete data analysis for oxidative and salt stress microarray experiments.

Begin microarray expression profiling of MR-1 cells exposed to toxic levels of certain metals (e.g., strontium).

Phenotype characterization and gene expression analysis of various mutant strains (e.g., ΔompR, ΔenvZ, ΔoxyR, Δdps).

Proteomics (Diversa)

Objectives:
The activities in this project are mainly in three areas: sample preparation, sample analysis, and data analysis.
Characterized the whole proteome of *Desulfovibrio vulgaris*
Establish methodology to quantify the change of protein expression upon cell stress

Progress since last report:
Data analysis for protein quantitation: To take a close look at the method of protein quantitation by ion intensity sum only, we tabulate the intensity sum of all peptides identified from the mixtures of 6 known proteins, analyzed using 1D capillary LC-MS/MS. The proteins were processed with standard protocol as outlined previously for *Desulfovibrio vulgaris* cell extracts. The intensity sums were further grouped under the proteins to arrive at the total intensity sum for the proteins as shown in Table 1. The net intensity sums after subtracting away the baseline sums are normalized by the number of residues of the proteins. The percents of deviation from the average are given on the last column. It's encouraged to see that the errors are less than 35%. Further more the intensity sum of trypsin come in at ~1% of the total intensity sum which is in good agreement of the enzyme-to-substrate ratio of 1:50 used to generate the protein digest. The results suggested a possibility of relative quantitation across different proteins.

Table 1. The normalized ion intensities calculated from the equal-molar mixtures of 6 known proteins.

Protein	# Peptide	# AA	Precursor Sum	Baseline Sum	Net	548793645	Protein	N.I.
ovalbumin	89	396	146551979700	937188830	145614790870	367714118	ovalbumin	-33.00
albumin	93	621	340608944680	1095920028	339513024652	546719846	albumin	-0.38
myoglobin	78	157	105326744600	848119320	104478625280	665468951	myoglobin	21.26
lactoglobulin	50	153	112464544900	555849380	111908695520	731429382	lactoglobulin	33.28
casein		229						
cytochrome c	29	104	45252782450	258645868	44994136582	432635929	cytochrome c	-21.17
trypsin	11	253	4529641100	100452857	4429188243	17506673	trypsin	-96.81

The next step is to look at the mixtures with a wider range of molar ratios. A new mixture was prepared so the concentrations range from 16 fmol to 10 pmol. The intensity sums were calculated as outlined. The correlation of the sums against the expected protein amounts is depicted in Figure 1.
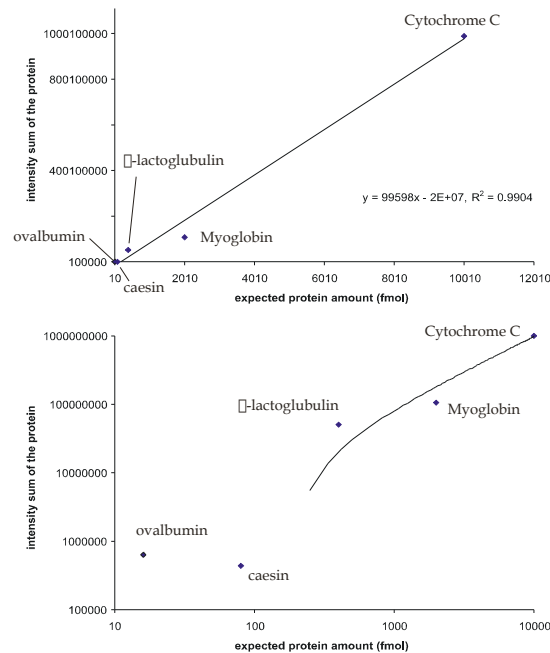
Figure 1. The correlation of the intensity sum vs the protein amount for several protein standards in the mixtures. The same data set is displayed as linear (top) or log-log plots (bottom). The solid lines in both plots are the linear fit of the data.

Very good linear correlation was observed ($R2 = 0.99$) as shown in the top plot. The trend was even more evident in the log-log view on the bottom. The protein at the lowest end of the concentration spectrum (BSA) was not detected.

The preliminary data is promising for absolute or relative protein quantitation based on ion intensity sums. The discrepancy in the peptide ionization efficiency may be compensated by the average of many peptides from the same proteins. In this regard, the enhanced sequence coverage offered by the 3D LC-MS/MS technique will be quite beneficial. Future works will look into the standard proteins in the presence of complex cell protein extracts to examine the extent of "matrix effect" on protein quantitation.

Protein complexes (Sandia)

Objective

> To work towards isolation of protein complexes involved in heat-shock, oxygen stress, and UV-radiation in *D. vulgaris* and their identification by separation (2D-gel/LC) followed by MS.
> Acquire mutants (and pertinent protocols) from collaborators to start designing affinity-tags.

Progress since last report
D. vulgaris

Replicated growth curves for cultures from Judy Wall's group as well as cultures from ATCC. Estimated average doubling time to be ~280 min for the former and ~192 min for the latter.

Performed Oxygen stress and pH stress runs under early-mid log phase conditions.

Extracted genomic DNA from both cultures for PCR runs to ascertain culture properties.

Obtained primers for NfrH (Nitrite reductase) and cytochrome c3 genes to be amplified.

Estimated whole cell protein extracts from heat shock experiments to be in the range of 3-7 mgs from 30 ml of starting cultures.

Performed western analysis of proteins extracted under heat shock conditions using mouse Hsp70 antibody and the Western Breeze kit (Invitrogen) - detected Hsp70 homolog under heat stress conditions.

Obtained C3 mutants (as well as wild-type) of G20 from Judy Wall for future isolation of protein complexes associated with cytochrome protein.

E. coli K12

Estimated whole cell protein extracts from UV radiation experiments to be in the range of 7-10 mgs from 50 ml of starting cultures.

Performed western analysis of proteins extracted under UV shock conditions using mouse RecA* antibody and the Western Breeze kit (Invitrogen) - detected about 2 fold higher expression of RecA under stress conditions.{*Description of Rec A: It is usually present in a bacterial cell at about 7200 molecules and can be easily induced 15-20 fold by simply exposing the cell to UV light, ionizing radiation or other DNA damaging agents. This level of induction makes it an ideal bait protein to pull down complexes in the absence of mutants and over-expression systems. Upon activation the RecA protein interacts with other gene products containing the SOS response elements in their promoter elements.}

Future Work

Identify proteins involved in heat-shock and oxygen stress by ICAT and 2D gels. The 2D gels (and DIGE) are already in progress. ICAT will be done in collaboration with Jay Keasling in his laboratory.

Mass Spectrometry for protein complex identification: Our immediate goal is to configure the QTOF. The existing electrospray source is inadequate for our samples and hence, we plan to install a nanospray ion source. This will give us the option to quickly analyze low volume (i.e. 10 microliters) samples, such as those obtained from in-gel protein digests, without LC/MS. Concurrently, we will setup a capillary LC system and develop Data Dependent Acquisition parameters for LC/MS on the QTOF. This will allow us to analyze more complex samples and to obtain data on lower abundant sample components than is possible by static nanospray. In addition, we will research the possibility of connecting an AP-MALDI source to our ion trap mass spectrometer. Having MALDI MS readily available to us will improve sample throughput and greatly aid in methods development for our initial experiments.

Obtain recombinant DNA protocols from collaborators (Judy Wall for *D. vulgaris* and Joe Zhou for *S. oneidensis*) to start designing affinity tags.

Metabolomics (UCB LBNL)

Objectives

The short-term goal is to develop a sensitive and quantitative LC/MS method to analyze nucleotides and their derivatives of *D. vulgaris* under various stresses. The metabolites we have been focusing are: NTP, NDP, NMP, dNTP, CoAs, NADP, NADH, NADPH, NAD, FAD, and various NTP/NDP-glucose.

To establish ICAT methodologies to be used for protein quantitation (in collaboration with Sandia who are looking at protein complexes).

Progress since last report

Product ions of nucleotides and derivatives have been determined. The common feature is that nucleotides containing same bases generate identical product ions. For example, ATP, ADP, AMP and dATP all produce same fragment (136 m/z) under the regular collision conditions. This feature allows us to apply precursor scan modes to selectively monitor a certain type nucleotides from complex samples. A method using ion-pair reverse phase HPLC is being developed. Two ion-pair reagents (tetrabutylammonium acetate and tetraethylammonium acetate) have been examined, but both significantly suppress ion signals and did not lead to the detection of corresponding nucleotides signals. Optimization of CoAs separation using reverse phase HPLC was performed, and slightly increased separation efficiency was achieved.

We constructed a degassing manifold for growing *Desulfovibrio* in our lab. An anaerobic atmosphere inside the anaerobic chamber was established and tested. Preliminary attempts at making growth media for *Desulfovibrio* were completed.

Towards the goal of comparing the microbial proteome prior to and after a stress, the ICAT methodology was established. A methods development protocol was conducted using known quantities of bovine serum albumin.

Future work

Two new ion-pair regent (tetrabutylammonium bromide and N,N-dimethyhexylamine) which are compatible with ionspray processes will be used to separate nucleotides and derivatives (standard mixture). MRM scan modes will be employed to develop a quantitative analysis of these samples. A method to extract metabolites of *D. vulgaris* will be developed based on our current protocol.

Successfully culture *Desulfovibrio* in our lab and establish oxygen stress conditions.

Data analysis of ICAT results using the ProICAT software and analysis of more complex protein mixtures (i.e. protein complexes and proteomes)

IV. Computational Core

During the past month, we focused much of our effort on restructuring the microbial genome database and API for the comparative genomics pipeline to accommodate the following requirements: (1) the database must interact quickly with the genome browser we are designing (all browse-able features can be queried by genomic position); (2) the database must support large amounts of partial and draft quality sequence; (3) the database must support the 600+ genomes expected to be available in the near future; (4) software tools to update the database must be parallelized to run on a cluster of computers. We have a trial version of the new database built and are currently testing it, we plan to finish the Perl-based interface in the next week. We have parallelized the database update utilities, and we hope to install a computer cluster on-site early next month.

Another main focus has been to bring genome-wide cis regulatory element prediction methods "in-house", with the predictions for Desulfovibrio vulgaris as the end goal. This includes phylogenetic footprinting, single genome motif-detection, motif clustering, and tying motif detection to the comparative bioinformatic tools we have already implemented (gene neighbors, phylogenetic profiling). As a test case, we have completed phylogenetic footprinting on 700 orthologous genes in three cyanobacterial species, and plan to cluster these motifs using two different published algorithms. From this test, we will learn which motif-finding algorithms and motif clustering algorithms perform best, and identify which currently unsequenced species would be the best target for a sequencing project aimed at uncovering regulatory motifs in Desulfovibrio vulgaris.

In addition to work on these two projects, we have produced some additional analysis/annotation of the Desulfovibrio vulgaris genomic sequence. In particular, we have new gene models based on CRITICA results, and have combined these with the TIGR models. We also have operon predictions using in-house methods based on comparative analysis. Both of these data sets have been posted to the biofiles website.

Over the next month, we plan to install a cluster of computers on-site and migrate to the database that we are currently testing. In the next month we will also begin work on a genome browser for the VIMSS site. We plan to complete our testing of cis-element prediction methods, and make a recommendation for a new sequencing project to the Joint Genome Institute. We will also hire and train a new computational biologist in the next month. Finally, we have begun to collect previously published microbial microarray data, and will begin to build a comprehensive in-house database of these experiments to compare against our own results for Desulfovibrio vulgaris.